

Grounded Knowledge Architecture & AI Dossier Pipelines

Turn messy enterprise archives into high-density, verified source dossiers engineered for Google NotebookLM™ and enterprise LLMs.  
garovartabed@gmail.com | 718-872-8578 | Austin, Texas | garovartabedian.com/services | Case Study

THE 'GARBAGE IN, GARBAGE OUT' BARRIER IN ENTERPRISE AI

Teams want to use generative tools and Google NotebookLM™ to synthesize company knowledge, but raw corporate files are messy: scanned PDFs, fragmented SOPs, contradictory spreadsheets, and disjointed manuals. Off-the-shelf AI hallucinates, drops critical nuance, and loses source context when fed uncurated files.

THE 4-STAGE KNOWLEDGE PIPELINE

|                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                                       |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <div><div>1. Ingestion &amp; Sanitization</div><div>Cleans OCR errors, extracts complex tables into readable Markdown, and strips formatting noise from proprietary PDFs, Word documents, and manuals.</div></div> | <div><div>2. Structural Chunking &amp; Metadata</div><div>Organizes raw documents into coherent modules stamped with structured YAML frontmatter (dates, authors, document version, access boundaries, and topic tags).</div></div>                                   |
| <div><div>3. Source-Linked Provenance</div><div>Connects dossier passages to supplied records with paragraph-level citation anchors so reviewers can check the source of a claim.</div></div>                      | <div><div>4. Target-Optimized Synthesis</div><div>Formats dossiers for Google NotebookLM™, Claude Projects, or private RAG, with source notes and questions for evaluating answers against supplied material. Generated outputs still need source checks.</div></div> |

LIVE ENGINEERING PROOF

An independent research project on garovartabedian.com demonstrates the approach with multilingual texts, commentary, and historical sources. Its Dossier Engine compiles curated research into downloadable, source-attributed Markdown for review before upload to NotebookLM.  
Inspect the live Dossier Engine: <https://garovartabedian.com/scriptures/dossier>

HOW WE WORK TOGETHER

|                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                              |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <div><div>Workflow &amp; Knowledge Diagnostic (\$500 fixed)</div><div>A bounded, one-week diagnostic of your document archives: identifying friction, hallucination risks, tool handoffs, and delivering a proof-of-concept sample dossier.</div></div> | <div><div>Knowledge Pipeline Sprint (Custom Scope)</div><div>A focused one-week sprint converting up to 50 complex corporate documents into an indexed, tagged, and verified source dossier library ready for NotebookLM or internal LLM search.</div></div> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|